一种基于动态加权 PPI 网络的关键蛋白质识别算法 *

杨书新，鲁纪华，汤达荣

(江西理工大学信息工程学院，江西 赣州 341000)

摘　要：与静态 PPI 网络相比，动态 PPI 网络更能体现蛋白质之间相互作用的真实情况，并有效降低 PPI 网络中的假阴性。现有的关键蛋白质预测方法主要应用在静态 PPI 网络，忽视了 PPI 网络的动态特性。为有效预测关键蛋白质，利用基因表达数据提取蛋白质的动态信息，再结合静态 PPI 网络构建动态 PPI 网络，然后引入 GO 术语对网络加权，并基于动态加权 PPI 网络提出一种新的预测方法-DWE。该方法以蛋白质在动态网络中的动态加权边之和与蛋白质在动态网络中出现的次数的比值衡量蛋白质在网络中的关键性。实验结果表明动态加权 PPI 网络有助于提高关键蛋白质的预测精度，且 DWE 方法优于其他几种关键蛋白质预测方法。

关键词：动态网络；关键蛋白质；GO 术语；动态加权 PPI 网络

中图分类号：TP301　　doi: 10.3969/j.issn.1001-3695.2017.08.0707

Novel algorithm prediction of essential proteins based on dynamic weighted PPI network

Yang Shuxin, Lu Jihua, Tang Darong

(*school of information engineering Jiangxi University of Science and Technology*, *Ganzhou* 341000, *China*)

Abstract: Compared with static PPI network, the dynamic PPI network can reflect the real situation of interactions among proteins and effectively reduce the false negation in PPI network. Most of the existing methods are based on static PPI network, which neglect the inherent dynamics of PPI network. To effectively predict the essential proteins, the dynamic information of proteins is extracted from gene expression data and integrate with static PPI network to construct dynamic PPI, then weighting the interaction between proteins based on GO term similarity and introduce a new method named DWE. This method assesses the important of one protein through the ratio of the sum of the dynamic weighted edge connecting this protein and the number of temporal networks that contain this protein The result shows that dynamic weighted PPI network help to improve the prediction accuracy of essential proteins and DWE outperforms other methods.

Key Words: dynamic network; essential proteins; GO terms; dynamic weighted PPI network

0　引言

关键蛋白质是生命活动的必需蛋白质。识别关键蛋白质不仅有助于理解生命活动，而且对药物设计及疾病治疗[1]具有重要意义。生物研究者通过生物实验预测关键蛋白质，但花费高、效率低，且只适用于部分生物体。因此，研究人员提出一系列的计算方法识别关键蛋白质，以提高预测效率，降低成本。

现有的关键蛋白质预测方法依据 PPI 网络的不同可以分为基于静态 PPI 网络的关键蛋白质识别和基于动态 PPI 网络的关键蛋白质识别。基于静态 PPI 网络的预测方法主要有度中心性(Degree Centrality, DC)[2]、介数中心性(Betweenness Centrality, BC)[3]、接近度中心性(Closeness Centrality, CC)[4]、子图中心性(Subgragh Centrality, SC)[5]、特征向量中心性 (Eigenvector Centrality, EC)[6]、信息中心性 (Information Centrality, IC)[7]、局部平均联通性(Local Average Connectivity, LAC)[8]和邻居中心性(Neighbor Centrality, NC)[9]等基于拓扑特征的预测方法，及融合生物信息的关键蛋白质预测方法，如 PeC[10]、DWC[11]、UDoNC[12]等。同时，2016 年，催鑫等[13]在已有的 CPPK 模型中引入 LDA 模型预测关键蛋白质。同年，杨莉萍等[14]结合边聚集系数和随机游走模型进行关键蛋白质识别。目前，对动态 PPI 网络的研究主要集中在蛋白质复合物和功能模块的挖掘，针对关键蛋白质检测方法的研究还不多。2015 年，Xiao 等[15]依据时间序列模型构建动态 PPI 网络，并将六种现有的关键蛋白质预测方法应用于动态 PPI 网络。与基于静态 PPI 网络的预测方法相比，该方法考虑了 PPI 网络的动态特性对预测结果的影响，但忽略了 PPI 网络中假阳性和假阴性的影响。为降低假阳性和假阴性的影响，本文在考虑 PPI 网络动态特性的基础上结合蛋白质功能相似性构建一种新的动态加权 PPI 网络（dynamic

基金项目：国家自然科学基金资助项目（61662028，41362015）；江西省教育厅重点科技项目（GJJ161566）

作者简介：杨书新（1978-），男，江西九江，副教授，博士，主要研究方向为图数据管理、生物信息学；鲁纪华（1989-），女，河南沈丘，硕士，主要研究方向为数据挖掘、生物信息学；汤达荣（1991-），男，江西赣州，硕士，主要研究方向为生物信息学.

weighted PPI network, DWPPIN）,并依据 DWPPIN 提出一种新的关键蛋白质检测方法-DWE。

1　基于动态加权 PPI 网络的关键蛋白质预测

基于动态加权 PPI 网络进行关键蛋白质预测的方法的具体流程：首先以静态 PPI 网络为框架，利用基因表达谱构建动态 PPI 网络；其次对动态 PPI 网络中的每个瞬态网络，利用边聚集系数和蛋白质之间的功能相似性衡量边的可靠性；最后计算动态加权网络中蛋白质的得分，得到蛋白质的重要程度。

1.1　动态 PPI 网络构建

动态 PPI 网络可通过蛋白质的共表达特性构建。蛋白质的共表达是指两个蛋白质之间存在相互作用，且在同一时刻的都属于活性蛋白质。表达水平超过阈值的蛋白质被认为是活性蛋白质。由于不同的蛋白质具有不同的表达水平，因此蛋白质的活性状态不能采用统一的阈值判断[16]。本文采用蛋白质的平均表达水平作为各自的阈值，即当蛋白质在某一时刻的表达值高于其平均表达水平时，本文认为蛋白质处于活性状态。

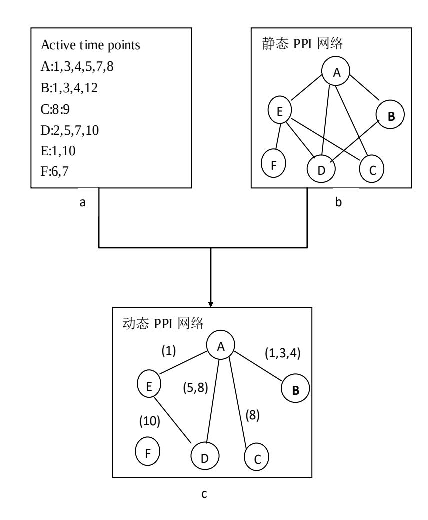


图 1　动态网络构建过程

基因表达谱包含各个蛋白质 3 个代谢周期(每个周期 12 个时刻，共 36 个时刻）的表达水平数据。每个蛋白质的平均表达水平可表示为：

$$\bar{T}_v=\frac{\sum_{i=1}^{36}t_v(i)}{36} \tag{1}$$

式（1）中，$t_v(i)$ 表达蛋白质 v 在 i 时刻的表达值，$\bar{T}_v$ 表示蛋白质 v 的平均表达值。

计算出各个蛋白质的平均表达水平后，依据基因表达谱判断各个蛋白质在每个时刻的活性状态，以确定各时刻 PPI 网络中的相互作用是否存在。对于静态 PPI 网络中的相互作用 $e(v,u)$，如果 v,u 在 i 时刻同时处于活性状态，那么 v,u 之间在 i 时刻出存在相互作用。可用公式表示为：

$$e_i(v,u)=\begin{cases}1 & t_v(i)>\bar{T}_v,t_u(i)>\bar{T}_u\\0 & otherwise\end{cases} \tag{2}$$

式（2）中，$e_i(v,u)=1$ 表示相互作用 $e(v,u)$ 在 i 时刻存在。反之，则不存在。

上述方法构建的动态 PPI 网络包含 36 个不同的瞬态 PPI 网络，每个瞬态网络都包含不同的活性蛋白质和不同的相互作用。动态 PPI 网络可形式化定义如下：

定义 1　动态 PPI 网络 DG=｛$G_1,G_2,\cdots,G_i,\cdots,G_l$｝，$l$ 表示时刻数，G_i=｛V_i,E_i｝是 i 时刻的 PPI 网络，$V_i=\{v_{i1},v_{i2},\cdots,v_{in}\}$ 表示 i 时刻处于活性状态的蛋白质集合，$E_i=\{e_{i1},e_{i2},\cdots,e_{im}\}$ 表示 i 时刻蛋白质相互作用的集合，$m=|E_i|$。

以图 1 为例说明动态网络的构建过程。静态 PPI 网络中 A、B 之间存在相互作用，且在 1、3、4 三个时刻同时处于活性状态。那么 A、B 之间的相互作用仅在 1、3、4 三个时刻出现。静态 PPI 网络中 E、C 之间存在相互作用，但在任一时刻，E、C 的活性状态都不相同。因此，在任一时刻，E、C 之间的相互作用都不存在。尽管 F、D 在时刻 7 同时处于活性状态，然而在静态 PPI 网络中，它们之间不存在相互作用，因此 F、D 之间在该时刻不存在相互作用。

1.2　动态 PPI 网络加权

基于对关键蛋白质的分析，关键蛋白质倾向于成簇出现，且蛋白质的关键性与相互作用的可靠程度之间关系密切。本文同时考虑蛋白质拓扑特征和生物特征，采用 GO 语义相似性和边聚集系数衡量相互作用的可靠程度。

相互作用的可靠程度可通过蛋白质之间的功能相似性评价。两个蛋白质之间的功能相似性越高，其相互作用越可靠。GO 术语是对基因的功能进行注释的结构化自然语言。研究证实，两个蛋白质之间的共有 GO 术语越多，它们的生物功能越相似。因此，蛋白质间相互作用的可靠性可通过它们在 GO 术语上的功能相似性进行评价。

令 GO（v）表示蛋白质 v 的 GO 注释信息组成的集合，GO_sim 表示两个蛋白质的 GO 语义相似性。那么两个蛋白质 v 和 u 之间的功能相似性计算公式如下：

$$GO_sim(v,u)=\frac{|GO(v)\cap GO(u)|}{|GO(v)\cup GO(u)|} \tag{3}$$

式（3）中，$GO(v)\cap GO(u)$ 表示 v 和 u 的共同 GO 术语集合。$GO(v)\cup GO(u)$ 表示存在于 v 或 u 的 GO 术语集合。

关键蛋白质不仅与相互作用的可靠性有关，而且往往成簇出现。边聚集系数作为 PPI 网络的一个重要拓扑特征，它不仅可以描述蛋白质之间相互作用的重要程度，还可评估蛋白质之间属于同一簇的概率，且能够较好的识别 PPI 网络中的关键蛋白质[17]。边聚集系数的计算公式如下：

$$ECC(v,u)=\frac{\tan(v,u)}{\min(d(v)-1,d(u)-1)} \tag{4}$$

式（4）中，$\tan(v,u)$ 表示 v 和 u 共同参与构成的三角形个数，$d(v)$ 和 $d(u)$ 分别表示 v 和 u 的邻居节点个数。

定义 2　动态加权 PPI 网络 DWG=$\{G_1,G_2,\cdots,G_l\}$，l 表示时刻数，$G_i=\{V_i,E_i,WE_i\}$ 是 i 时刻的加权 PPI 网络，

$V_i=\{v_{i1},v_{i2},\cdots,v_{in}\}$ 表示 i 时刻处于活性状态的蛋白质集合，$E_i=\{e_{i1},e_{i2},\cdots,e_{im}\}$ 表示 i 时刻蛋白质相互作用集合，$WE_i=\{we_{i1},we_{i2},\cdots,we_{im}\}$ 是相互作用权值的集合，$m=|E_i|$。

动态 PPI 网络中个瞬态网络的相互作用权值计算公式如下：

$$WE(v,u)=GO_sim(v,u)+ECC(v,u) \tag{5}$$

1.3 关键蛋白质识别

蛋白质的关键性可根据蛋白质在 2.2 节构建的动态加权 PPI 网络中的评分判定。由于瞬态网络中的活性蛋白质都各不相同，所以在计算蛋白质的评分时会将其在动态加权 PPI 网络中出现的次数考虑在内。蛋白质的评分 DWE 为各个瞬态网络中该蛋白质与其邻居节点之间权值之和与其在动态网络中出现的次数之比。其计算公式为：

$$DWE(v)=\frac{\sum_{i=1}^{36}\sum_{u\in N_i(v)}WE_i(v,u)}{fre(v)} \tag{6}$$

式（6）中，$N_i(v)$ 表示在 i 时刻对应的瞬态网络中节点 v 的邻居节点集合；$WE_i(v,u)$ 表示在 i 时刻对应的瞬态网络节点 v 和节点 u 之间边的权值；$fre(v)$ 表示动态网络中包含节点 v 的瞬态网络的个数，即节点 v 在动态网络中出现的次数。

本文根据每个蛋白质的 DWE 值，将所有的蛋白质按照 DWE 值进行降序排列。DWE 值越高的蛋白质，排名越靠前，越可能是关键蛋白质。将排序后的前 K 个蛋白质输出作为 DWE 方法识别出的候选关键蛋白质。K 是动态加权 PPI 网络中包含关键蛋白质的个数。DWE 方法的伪代码描述如下：

```
DWE 算法:
Input: Static PPI network G, gene expression data GED, GO annotation GOA, K
Output: the top K Ranked proteins by DWE value in descent order
Construct dynamic PPI network DG={G_1,G_2,...,G_i} based on GED according to the formula(1) and (2)
For each G_i ∈ DG do
  For each e_ij(v,u) ∈ E_i do
    compute GO_sim(v,u)
    compute ECC(v,u)
    compute WE(v,u)
  end for
  for each v_ij ∈ V_i do
    compute the sum of WE(v_ij,u)
  end for
end for
a) Compute the DWE of all nodes by formula（6）
b) Return the top K ranked proteins according to their DWE values
```

DWE 算法第 1）步根据蛋白质的基因表达数据和静态 PPI 网络构建动态 PPI 网络。第 2）步实现动态 PPI 网络的加权，对于动态 PPI 网络中的任意瞬态网络 G_i，算法循环的以式（5）计算每条边的权值，然后对瞬态网络 G_i 中任意节点 v_{ij}，计算与节点 v_{ij} 相连边的权值之和。算法第 3）步对 PPI 网络中的所有蛋白质评估其在动态加权 PPI 网络中的 DWE 评分，并按照 DWE 值将蛋白质进行降序排列。在 4）步返回排序后的前 K 个蛋白质作为 DWE 算法预测的 K 个关键蛋白质。

2 实验结果与分析

2.1 实验数据

由于酵母蛋白质相互作用网络和关键蛋白质数据都具有较高的完整性和可靠性，因此本文选择酵母蛋白质相关数据集验证本文方法的有效性。实验用到的相关数据描述如下：

（1）酵母蛋白质交互作用网络（Protein-Protein Interaction, PPI）来自 DIP 数据库[18]。其中包含 5093 个蛋白质，24743 个相互作用。

（2）标准关键蛋白质数据可通过整合 MIPS[19]、SGD[20]、DEG[21]、SGDP[22]4 个数据库的数据得到，包含关键蛋白质 1285 个，其中出现在酵母 PPI 网络中的仅有 1167 个。

（3）酵母基因表达谱来源于 NCBI 基因表达综合网站[23]编号为 GSE3431 的数据。它以一个 $m\times n$ 矩阵的形式存储，矩阵中的每一个值表示某个蛋白质在某个时刻的表达水平。基因表达谱包含 6740 个基因产物的表达水平数据。其中有 4981 个基因产物出现在酵母 PPI 网络中。

（4）酵母蛋白质 GO 注释信息下载自基因本体数据库[24]（2016 年 12 月 24 日的版本），它主要包括三部分：生物过程、分子组件和分子功能。

2.2 实验结果

为验证动态加权 PPI 网络的效果和评估 DWE 方法的识别效率，本文设计了两组实验进行对比分析。其中一组使用三种不同的 PPI 网络数据，包括静态 PPI 网络（Static PPI Network, SPPIN）、Xiao 等[13]构建的动态 PPI 网络（Active PPI Network, APPIN）及本文构建的动态加权 PPI 网络（Dynamic Weighted PPI Network, DWPPIN）。并在每种 PPI 网络上，分别应用中心性方法预测关键蛋白质。另一组编程实现现有的七种经典的中心性检测方法、融合了生物信息的 PeC、WDC 和 UDoNC 方法然后与本文提出的 DWE 方法进行对比。并采用“排序-筛选”法、“刀切法”及敏感度、特异性等指标评估实验结果。

2.2.1 PPI 网络对检测效果的影响

由于 DC、LAC、NC 分别从邻居个数、邻居的重要性和边的重要性三个不同的角度衡量蛋白质在网络中的重要性，因此选择这三种方法作为关键蛋白质检测方法。图 2 给出了这三种方法在不同 PPI 网络中识别关键蛋白质的数量。从图 2 可看出，与静态 PPI 网络相比，动态加权网络的检测结果明显提高了很多。当选取前 600 个作为候选关键蛋白质时，动态加权 PPI 网络的正确率提高了 20%。尽管在候选关键蛋白质较少时，动态加权 PPI 网络和动态 PPI 网络的识别正确率非常接近，然而随

着候选关键蛋白质的增加，动态加权 PPI 网络的识别正确率逐渐优于动态 PPI 网络。因此动态加权 PPI 网络能够提高关键蛋白质的检测效果。

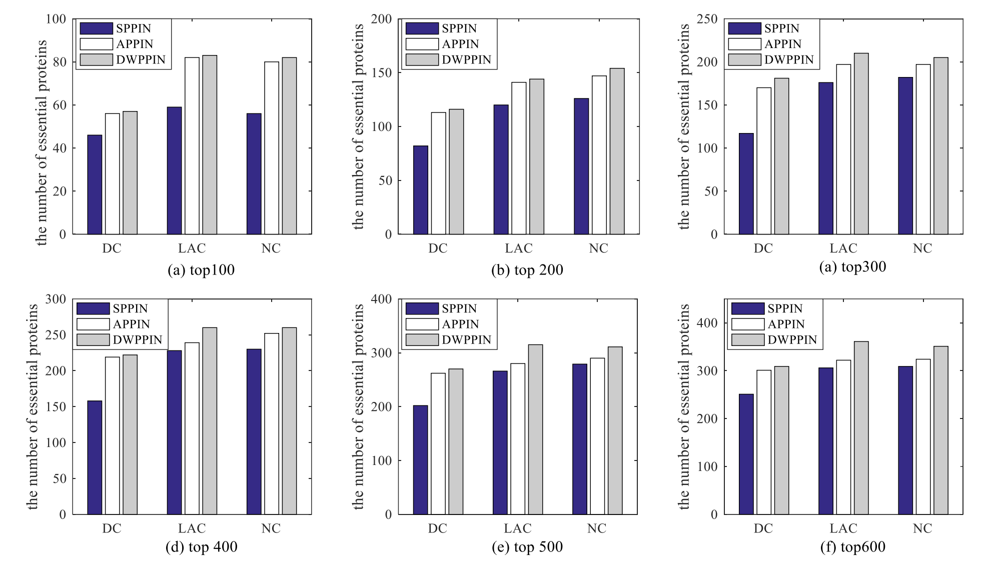


图 2 在三种不同的 PPI 网络中三种方法识别关键蛋白质的数量

2.2.2 “排序-筛选”法分析

“排序-筛选” 法的具体做法是：对所选的测度参数，计算蛋白质网络中每个蛋白质节点的相应测度值，并将蛋白质节点按照测度值由大到小进行排序，选取排序靠前的蛋白质作为预测的关键蛋白质，然后与已知的关键蛋白质数据集比对，从而得到该测度参数识别出的关键蛋白质数目。本实验中，分别选取排序在前 1%、5%、10%、15%、20%、25%的蛋白质作为候选关键蛋白质。

图 3 为 11 种方法的关键蛋白质识别结果。从图 3 可看出，七种经典的中心性方法中的 NC 识别效果最好。与 NC 相比，DWE 的识别正确率分别提高了 28.12%、15.72%、15.90%、11.52%、8.41%、6.62%。同时，与融合生物信息的识别方法 PeC、WDC、UDoNC 相比，当候选关键蛋白质不超过 25%时，DWE 仍然有比较明显的优势。虽然随着候选关键蛋白质的增加，DWE 的优势逐渐减弱，但当选取 25%的蛋白质为候选关键蛋白质时，DWE 的识别正确率仍然有所提高。这时，DWE 的识别正确率提高了 8.21%、2.47%、1.22%相比于 PeC、WDC、UDoNC。

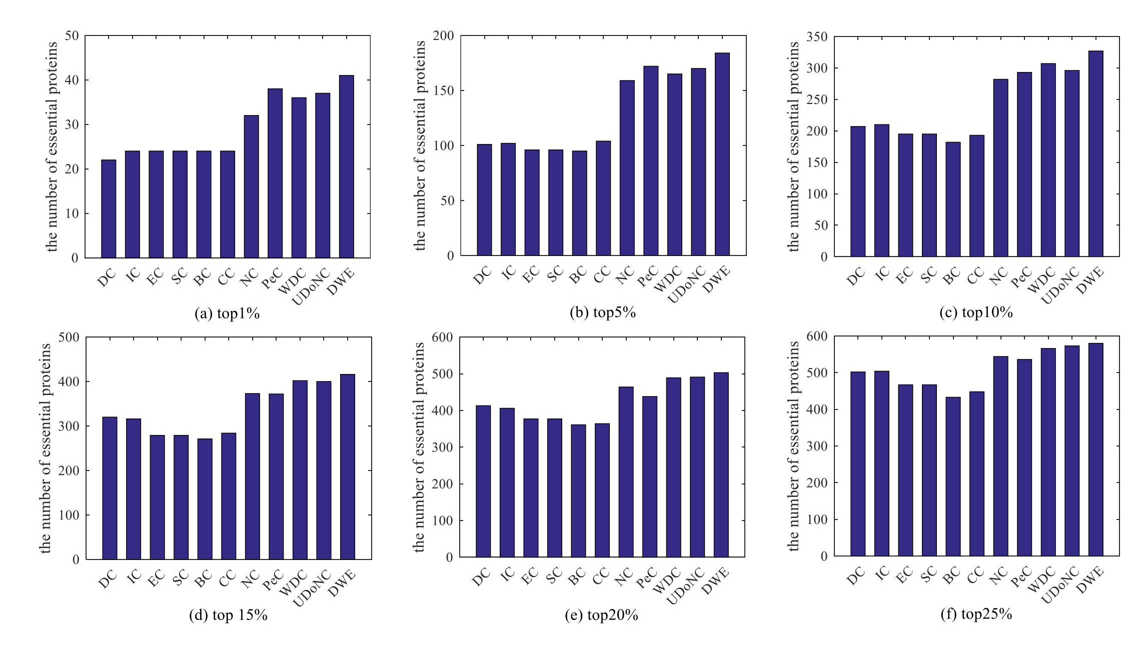


图 3 11 种方法在不同样本容量中预测的关键蛋白质数目

2.2.3 其他评估方法分析

为更细致的评估 DWE 方法的性能，本文引入“刀切法”对比分析 DWE 方法与其他十种中心性测度方法的预测结果。图 4 通过刀切法列出了各方法关键蛋白质的预测结果。图 4 中，X 轴表示累计候选关键蛋白质数量，Y 轴表示累计真实关键蛋白质数量。刀切法曲线下方的面积用于对比各个方法的性能，

面积越大说明识别率越高。从图 4 可看出，与 DC、BC、EC、SC、IC、NC 和 CC 相比，DWE 具有最好的识别率。从图 4（a）看出,DWE 曲线明显位于 PeC 曲线的上方，这说明 DWE 的识别率优于 PeC。同时从图 4（b）看出，DWE 的识别率高于 WDC 和 UDoNC 两种方法。

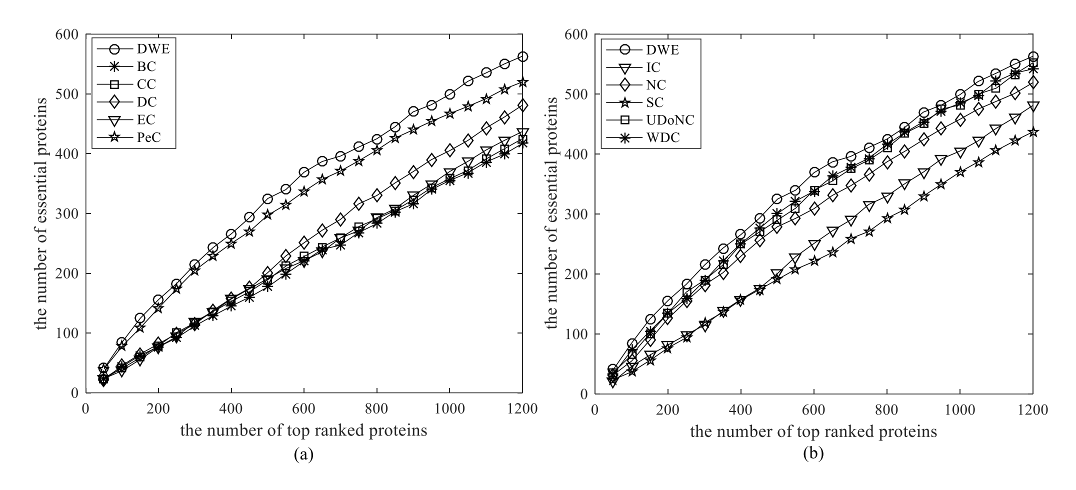


图 4　刀切法曲线图

此外，本文采用敏感度（SN）、特异性（SP）、F 测度（F-measure）、正确率（ACC）、阳性预测值（PPV）和阴性预测值（NPV）这六个评估指标对各个方法的性能进行评估。其计算公式如下：

$$SN = \frac{TP}{TP + FN} \tag{7}$$

$$SP = \frac{TN}{TN + FP} \tag{8}$$

$$PPV = \frac{TP}{TP + FP} \tag{9}$$

$$NPV = \frac{TN}{TN + FN} \tag{10}$$

$$F = \frac{2 \times SN \times PPV}{SN + PPV} \tag{11}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

式（7）-（12）中，TP 表示被预测为关键蛋白质的关键蛋白质数量；TN 表示被预测非关键蛋白质的非关键蛋白质数量；FP 表示被预测为关键蛋白质的非关键蛋白质数量；FN 表示被预测非关键蛋白质的关键蛋白质数量。

根据已知关键蛋白质的数量，本文选取排序后的前 1167 个蛋白质作为预测关键蛋白质，并根据已知的关键蛋白质计算各方法六项指标的值。表 1 为 DWE 与其他十种方法的六项指标的值。从表 1 可看出，DWE 的任一指标都高于其他几种方法。

3　结束语

本文利用 PPI 网络的动态特性，结合蛋白质的生物特征构建动态加权 PPI 网络，并提出一种名为 DWE 的关键蛋白质识别方法。该方法降低了数据噪声对实验结果的影响，提高了 PPI 网络的可靠性。与其他方法相比，DWE 方法提高了识别准确率，能够识别出更多的关键蛋白质。

表 1　DWE 和其他几种中心性测度的各个评价指标的比较

Methods	SN	SP	PPV	NPV	F-measure	ACC
DC	0.4002	0.8217	0.4002	0.8217	0.4002	0.7251
IC	0.4002	0.8217	0.4002	0.8217	0.4002	0.7251
EC	0.3676	0.8120	0.3676	0.8120	0.3676	0.7102
SC	0.3676	0.8120	0.3676	0.8120	0.3676	0.7102
BC	0.3505	0.8069	0.3505	0.8069	0.3505	0.7023
CC	0.3548	0.8082	0.3548	0.8082	0.3548	0.7043
NC	0.4353	0.8321	0.4353	0.8321	0.4353	0.7412
PeC	0.4362	0.8324	0.4362	0.8324	0.4362	0.7416
WDC	0.4619	0.8400	0.4619	0.8400	0.4619	0.7534
UDoNC	0.4653	0.8411	0.4653	0.8411	0.4653	0.7550
DWE	0.4685	0.8441	0.4685	0.8441	0.4685	0.7594

参考文献：

[1] Jeong H, Oltvai Z N, Barabási A-L. Prediction of protein essentiality based on genomic data [J]. ComPlexUs, 2002, 1 (1): 19-28.

[2] Jeong H, Mason S P, Barabási A-L, et al. Lethality and centrality in protein networks [J]. Nature, 2001, 411 (6833): 41-42.

[3] Joy M P, Brock A, Ingber D E, et al. High-betweenness proteins in the yeast protein interaction network [J]. BioMed Research International, 2005, 2005 (2): 96-103.

[4] Wuchty S, Stadler P F. Centers of complex networks [J]. Journal of Theoretical Biology, 2003, 223 (1): 45-53.

[5] Estrada E, Rodriguez-Velazquez J A. Subgraph centrality in complex networks [J]. Physical Review E, 2005, 71 (5): 056103.

[6] Bonacich P. Power and centrality: A family of measures [J]. American journal of sociology, 1987: 1170-1182.

[7] Stephenson K, Zelen M. Rethinking centrality: Methods and examples [J]. Social Networks, 1989, 11 (1): 1-37.

[8] Li Min, Wang Jianxin, Chen Xiang, et al. A local average connectivity-based method for identifying essential proteins from the network level [J]. Computational Biology & Chemistry, 2011, 35 (3): 143-50.

[9] Wang Jian, Li Min, Wang Huan, et al. Identification of Essential Proteins Based on Edge Clustering Coefficient [J]. IEEE//ACM Transactions on Computational Biology & Bioinformatics, 2012, 9 (4): 1070.

[10] Li Min, Zhang Hanhui, Wang Jianxin, et al. A new essential protein discovery method based on the integration of protein-protein interaction and gene expression data [J]. BMC Systems Biology, 2012, 6 (1): : 15.

[11] Tang Xiwei, Wang Jianxin, Zhong Jiancheng, et al. Predicting Essential Proteins Based on Weighted Degree Centrality [J]. IEEE//ACM Trans Comput Biol Bioinform, 2014, 11 (2): 407-18.

[12] Peng Wei, Wang Jianxin, Cheng Yingjiao, et al. UDoNC: An Algorithm for Identifying Essential Proteins Based on Protein Domains and Protein-Protein Interaction Networks [J]. IEEE//ACM Trans Comput Biol Bioinform, 2015, 12 (2): 276-88.

[13] Xiao Qianghua, Wang Jianxin, Peng Xiaoqing, et al. Identifying essential proteins from active PPI networks constructed with dynamic gene expression [J]. BMC Genomics, 2015, 16 (3): 1.

[14] 崔鑫, 邵明玉. 结合主题特征和互作用网络拓扑特性的关键蛋白质识别 [J]. 计算机应用与软件, 2016, 33 (08): 283-288.

[15] 杨莉萍, 路松峰, 黄钰. 一种基于随机游走模型的关键蛋白质预测方法 [J]. 华中农业大学学报, 2016, 35 (6): 86-91.

[16] Wang Jianxin, Peng Xiaoqing, Li Min, et al. Construction and application of dynamic protein interaction network based on time course gene expression data [J]. Proteomics, 2013, 13 (2): 301-312.

[17] Wang Huan, Li Min, Wang Jianxin, et al. A New Method for Identifying Essential Proteins Based on Edge Clustering Coefficient [J]. Lecture Notes in Computer Science, 2011, 6674: 87-98.

[18] Xenarios I, Salwínski Ł, Duan X J, et al. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions [J]. Nucleic Acids Research, 2002, 30 (1): 303.

[19] Hw M, D F, Kf M, et al. MIPS: analysis and annotation of proteins from whole genomes [J]. Nucleic Acids Research, 2006, 34 (2): 169-72.

[20] Cherry J M, Adler C, Ball C, et al. SGD: Saccharomyces Genome Database [J]. Nucleic Acids Research, 1998, 26 (1): 73-9.

[21] Zhang Ren, Lin Yan. DEG 5. 0, a database of essential genes in both prokaryotes and eukaryotes [J]. Nucleic Acids Research, 2009, 37 (Database issue): D455-D458.

[22] Saccharomyces Genome Deletion Project [EB/OL]. http://www-sequence. stanford. edu/group/yeast_deletion_project.

[23] Tu B P, Kudlicki A, Rowicka M, et al. Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes [J]. Science, 2005, 310 (5751): 1152-1158.

[24] Consortium T G O. Gene Ontology Consortium: going forward [J]. Nucleic Acids Research, 2015, 43 (Database issue): 1049-56.